Alzoubi and Mishra
EURASIP Journal on Information Security (2025) 2025:18
https://doi.org/10.1186/s13635-025-00203-9

EURASIP Journal on
Information Security

REVIEW

Open Access

Differential privacy and artificial intelligence: potentials, challenges, and future avenues

Yehia Ibrahim Alzoubi[1] and Alok Mishra[2*]

Abstract

Privacy preservation has become an increasingly critical concern in applications where data serves as a cornerstone for decision-making and innovation. Researchers and developers are dedicated to identifying and mitigating emerging risks while improving the privacy of existing systems. Artificial intelligence technologies can dynamically detect and address privacy concerns. Differential privacy, with its strong and verifiable assurances, is critical for addressing rising concerns about data privacy in the age of big data and advanced analytics. Combining differential privacy with AI has been identified as a solution for balancing data usage for insights while maintaining individual privacy. However, research in this field is still scarce due to the recent widespread application of artificial intelligence in many industries. This paper reviews current literature, professional websites, and other online resources to determine the potential, challenges, and future directions of combining differential privacy with AI. The key opportunities identified in this study include enhancing privacy (reported in 27% of the reviewed papers), promoting responsible AI (21%), facilitating data sharing (14.5%), and minimizing AI model biases (12.5%). Several concerns, however, require additional exploration, including accuracy trade-offs, computational complexity, regulatory restrictions, expertise, data usability, scalability constraints, and bias concerns. Given that this combination is still a relatively new field, AI developers and users need to stay current on differential privacy research and implement appropriate measures.

Keywords Artificial intelligence, Differential privacy, Privacy-preserving, Anonymization, Data protection

1 Introduction

Privacy preservation has become an increasingly critical concern in applications where data serve as a cornerstone for decision-making and innovation [1, 2]. With massive volumes of personal information being gathered, stored, and analyzed, the potential for data breaches and illegal access has increased, posing serious concerns about individual privacy. The need to improve privacy protection has never been greater [3]. Artificial intelligence (AI) solutions can dynamically detect and mitigate privacy issues, automate data anonymization, and monitor for possible breaches, encouraging a safe data environment [4]. Differential privacy is a privacy-preserving technique that ensures the confidentiality of individual data within a dataset by introducing random noise [5]. The importance of differential privacy lies in its robust and quantifiable privacy guarantees, which address the growing concerns about data privacy in the era of big data and advanced analytics [6].

AI comprises a wide array of technologies attempting to imitate human intelligence through robots [7]. Machine learning algorithms, natural language processing (NLP), and neural networks are essential AI techniques that allow systems to learn from data, interpret human language, and make judgments [8]. AI has numerous applications, including healthcare, where it aids in diagnostics and personalized medicine; finance; fraud detection; and algorithmic trading. Emerging generative AI, like OpenAI's GPT models, marks a big step

*Correspondence:
Alok Mishra
alok.mishra@ntnu.no
[1] College of Business Administration, American University of the Middle East, Egaila, Kuwait
[2] Faculty of Engineering, Norwegian University of Science and Technology (NTNU), Trondheim, Norway

forward, capable of producing new content ranging from text and graphics to music [9]. These models transform companies by automating creative processes, improving consumer experiences, and proposing novel solutions to complicated issues.

The combination of differential privacy and AI has been recognized as critical to tackling the twin issue of using data for insights while protecting individual privacy [10]. As the volume and sensitivity of data increase, so does the demand for powerful privacy-preserving approaches [11]. Differential privacy, which adds noise to datasets to safeguard individual identities, complements AI's data-driven capabilities by guaranteeing that personal information is kept private even while AI models extract useful patterns and insights [12]. Evolving research on this subject focuses on improving the integration of various technologies to create AI systems that are both powerful and privacy-conscious. This research is crucial for sectors like healthcare, finance, and social media, where data privacy is paramount, enabling the deployment of AI solutions that respect privacy regulations and foster public trust. Hence, the objective of our study is to explore the intersection of differential privacy and AI by examining existing literature, professional websites, and other online resources to uncover the potential, challenges, and future directions of this combination. The paper aims to address the following research questions:

RQ1: What are the potential benefits of integrating differential privacy with AI?

RQ2 What challenges are currently hindering the effective combination of these technologies?

RQ3: What future directions are necessary to fully realize the effective combination of these technologies?

The novelty of this study lies in its exploration of the integration of differential privacy with AI, a field that, to the best of the authors' knowledge, has received limited attention in existing research. Given the emerging nature of this domain, this paper establishes foundational guidelines by outlining the key potentials, challenges, and future directions for leveraging differential privacy in AI systems. Our study contributes to advancing knowledge in the field as follows:

- This paper conducts an extensive literature review of existing studies, professional websites, and online resources on differential privacy in AI to synthesize knowledge on differential privacy with AI integration.
- The paper identifies and analyzes the key benefits of differential privacy in AI, including enhanced data privacy and improved AI trustworthiness. The key benefits identified in this paper include enhancing privacy, promoting responsible AI, facilitating data sharing, and minimizing AI model biases.
- It addresses technical, ethical, and regulatory challenges that hinder effective differential privacy with AI integration including accuracy trade-offs, computational complexity, regulatory restrictions, expertise, data usability, scalability constraints, and bias concerns.
- It proposes future research directions such as developing robust privacy-preserving AI algorithms and exploring new applications. It also recommends advancements in evaluating trade-offs between data utility and privacy protection.

The rest of the paper is organized as follows: Sect. 2 delves into the research background of AI and differential privacy. Section 3 outlines the methodology employed in this study. Section 4 analyzes the potential of integrating differential privacy with AI. Section 5 addresses the challenges that hinder this combination. Section 6 discusses the implications of the findings, limitations, and future research directions in this area. Finally, Sect. 7 concludes the paper.

2 Background

Combining AI with differential privacy creates a powerful method for using extensive data while ensuring individual privacy. This section provides background on AI and differential privacy. It also discusses the related survey research conducted in this context.

2.1 Differential privacy

Differential privacy is a mathematical framework that protects individuals' privacy while allowing for data analysis [13]. It works by inserting properly calibrated noise into the data or the results of data queries, guaranteeing that the inclusion or absence of a single data point has no meaningful effect on the answer [14]. This method ensures that it is hard to deduce any single individual's data from the aggregated findings, offering high privacy protection [15]. Differential privacy is important where sensitive data, like medical records or financial transactions, must be examined for insights while maintaining the confidentiality of the persons involved [10]. Figure 1 illustrates a system incorporating differential privacy to protect sensitive user data while enabling data analysis. Sensitive user data is collected from various sources (e.g., education centers and clinics). The original data is transformed by adding noise using a specific function (the noise function) and considering the sensitivity of the data. This process aims to protect individual privacy while preserving data utility. The noisy data is processed using techniques like deep learning to extract

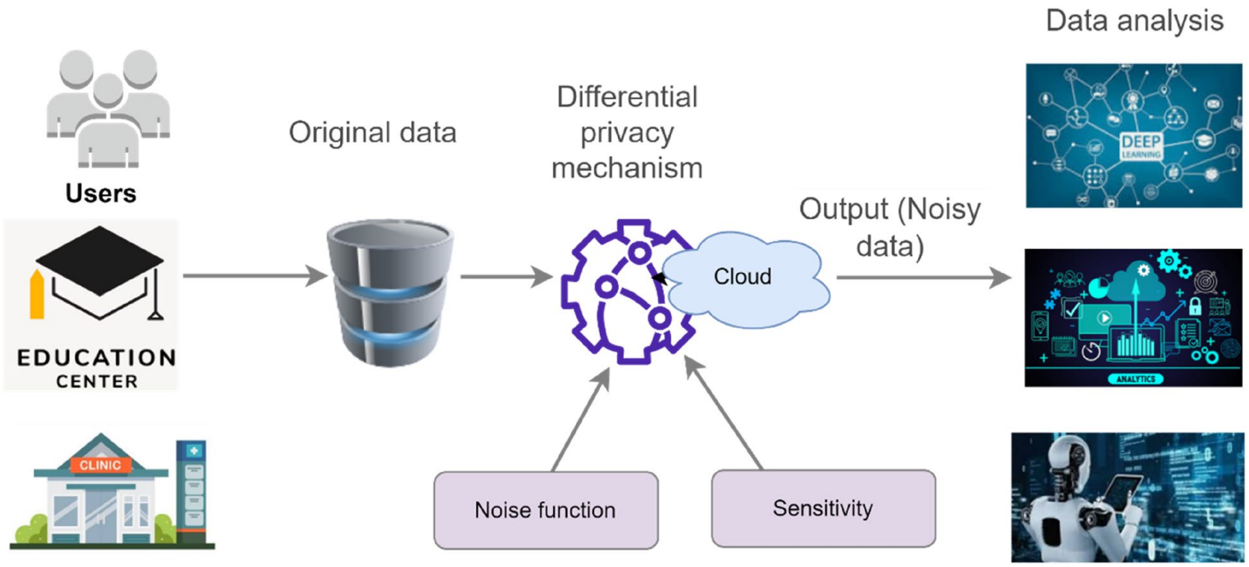

Fig. 1 Differential privacy concept

valuable insights without compromising an individual's privacy.

The significance of differential privacy stems from its strong and measurable privacy assurances, which answer rising concerns about data privacy in the age of big data and sophisticated analytics [16]. As businesses increasingly rely on massive datasets to drive decision-making and innovation, protecting individuals' privacy inside these databases becomes crucial [17]. Differential privacy protects data privacy by reducing the chance of re-identification or sensitive information exposure, even if an attacker has access to additional information [5]. This is critical for preserving trust between data suppliers and consumers, as well as adhering to strict privacy requirements like the general data protection regulation [18].

Differential privacy's strategy and assurances set it apart from conventional privacy techniques, such as encryption or anonymization [19]. While anonymization aims to eliminate personally identifying information from datasets, it is frequently susceptible to re-identification attacks when paired with other data sources [20]. Encryption, on the other hand, safeguards data during transmission and storage but does not prevent privacy violations during data analysis. Differential privacy bridges these gaps by ensuring that the analytical findings do not jeopardize individual privacy [21]. This makes it a valuable tool for balancing data value and privacy, distinguishing it from typical privacy-preserving strategies [22]. Major companies utilize differential privacy in various ways. Apple employs this method to gather anonymous usage insights from devices such as iPhones and Macs [11]. Facebook uses differential privacy to collect behavioral data for targeted advertising campaigns. Similarly, Amazon relies on this technique to understand personalized shopping preferences while protecting sensitive information [11].

Currently, no law explicitly mandates the use of differential privacy. However, differential privacy is often recommended as a privacy-preserving technique to comply with these laws, especially for handling sensitive data in AI and big data analytics. Here, we present an overview of the common acts that can be used in regulating differential privacy [23–25].

- General data protection regulation (GDPR): The European Union's (EU) comprehensive data privacy law came into effect in May 2018. It governs how organizations collect, process, and store the personal data of individuals within the EU. GDPR requires data minimization and privacy-by-design, which differential privacy can help achieve by ensuring data remains anonymized and non-identifiable.
- Health Insurance Portability and Accountability Act (HIPAA): It is a U.S. law enacted in 1996 to protect sensitive healthcare data. It applies to healthcare providers, insurers, and business associates handling protected health information (PHI). Differential privacy can be used to de-identify patient data while allowing healthcare research.
- California Consumer Privacy Act (CCPA): It was enacted in 2020 and enhances data privacy rights for California residents. It gives individuals greater transparency and control over how businesses collect, use, and share personal information. CCPA gives con-

sumers the right to opt out of data collection; differential privacy can help organizations provide aggregate insights while preserving individual privacy.

- Federal Trade Commission (FTC): It is a U.S. regulatory agency that enforces consumer protection and competition laws, including data privacy and security regulations. While it does not have a single overarching privacy law like GDPR, HIPAA, or CCPA, the FTC regulates privacy through various sector-specific laws and enforcement actions under the FTC Act (1914). FTC enforces privacy through consumer protection laws, and differential privacy can be used as a best practice to prevent companies from mishandling user data.

2.2 Artificial intelligence

AI covers several methods and technologies designed to allow machines to accomplish jobs that would normally require human intellect [26]. These tasks include data-driven learning, pattern recognition, decision-making, and NLP. AI approaches are typically classified into two types: machine learning, in which algorithms improve with experience, and deep learning, a subset of machine learning that consists of neural networks with several layers [26]. AI applications are used in several fields, such as healthcare diagnostics, fraud detection, and transportation. As AI evolves, it promises to significantly enhance innovation and productivity across many sectors [12].

Generative AI is a subfield of AI that focuses on synthesizing new content from existing data. Unlike standard AI, which can only categorize or forecast based on incoming data, generative AI can create new text, graphics, music, and more [9]. A notable example is ChatGPT, which produces human-like text. These technologies are transforming the creative industries by automating content generation and allowing new modes of expression. However, generative AI raises questions about authenticity and the possibility of misuse, such as deepfakes, which may mislead and manipulate the public [27]. Despite these obstacles, generative AI has enormous creative potential, opening up hitherto untapped avenues of creativity.

The link between AI and privacy is nuanced and diverse. AI systems frequently require vast volumes of data to work efficiently; thus, there is rising worry about how this data is acquired, kept, and used. Differential privacy can be considered an important solution for addressing these problems, allowing data to be utilized in AI models while protecting individual privacy [14, 28]. This technology introduces controlled noise into the data, guaranteeing that the result does not expose personal information about any individual [12]. As AI becomes more interwoven into everyday life, it is critical to strike a balance between its benefits and the need to preserve personal privacy [18]. Robust protections for privacy are necessary to develop confidence and ensure that the deployment of AI technology maintains user privacy and conforms with regulatory standards [9].

2.3 AI settings: federated learning vs. centralized learning and the role of differential privacy

In centralized learning, all data is collected and stored on a central server, where the AI model is trained. While this traditional approach benefits from direct access to large datasets, it introduces significant privacy risks including privacy issues because all raw data is stored centrally, and unauthorized access or data breaches can expose sensitive user information. Moreover, the role of differential privacy is to mitigate these risks by adding controlled noise to the data before or during the model training process, ensuring individual data points cannot be distinguished. Differential privacy ensures that the probability of obtaining a particular output *M(D)* from dataset D remains nearly the same even if a single record is removed or changed, where *M(D)* is the mechanism applied to dataset D, D' is a neighbor dataset differing by one entry, S is a subset of possible outputs, ϵ (epsilon) is the privacy budget, which controls the level of privacy (smaller ϵ means stronger privacy) [16].

$$P[M(D)\epsilon S] \leq e^{\epsilon} P[M(D\prime)\epsilon S] \tag{1}$$

The noise is often introduced using Laplace or Gaussian mechanisms. Laplace mechanism adds noise from a Laplace distribution, according to Eq. 2, where Δf is the sensitivity of the function [16]:

$$\text{Noise} \sim \text{Lablace}\left(0, \frac{\Delta f}{\epsilon}\right) \tag{2}$$

On the other hand, federated learning is a decentralized approach where multiple devices or institutions train a shared model collaboratively without sharing raw data. Instead, only model updates (gradients) are transmitted. Even though raw data isn't shared, gradients can still leak private information as malicious participants may infer sensitive details from model updates. Differential privacy in federated learning enhances privacy by adding noise to model updates before sending them to the central server. This prevents individual users'contributions from being extracted while still allowing the model to learn meaningful patterns. A commonly used differential privacy mechanism in federated learning is Differentially Private Stochastic Gradient Descent (DP-SGD), where Clipping represents the gradients of each user gi which are clipped to a fixed norm C, according to Eq. 3 [16].

$$\widetilde{g}i = \frac{gi}{\max(1, \parallel gi \parallel /C)} \tag{3}$$

The sensitivity Δf measures how much a function's output can change by modifying a single input, according to Eq. 4, where D and D' differ by at most one record [16]:

$$\Delta f = max_{D,D'} \parallel f(D) - f\left(D'\right) \parallel \tag{4}$$

For a function *f(D)*, the Laplace mechanism adds noise sampled from a Laplace distribution, according to Eq. 5, where *Lap(b)* is the Laplace distribution with scale $b{=}\Delta f/\epsilon$. Larger ϵ allows less noise, leading to lower privacy but better utility [16].

$$M(D) = f(D) + Lab\left(\frac{\Delta f}{\epsilon}\right) \tag{5}$$

2.4 Common machine learning techniques used in differential privacy

Various machine learning techniques can be integrated into differential privacy frameworks to enable data analysis while ensuring the protection of individual privacy. This section provides an overview of the most commonly reported techniques in the literature. Additionally, we explore how each technique can be applied in the context of differentially private learning, along with a comparative analysis of their respective advantages and limitations [29–31]. Table 1 summarizes these techniques with their weaknesses and strengths.

- NLP: NLP models process sensitive text data (e.g., medical records, user-generated content) while ensuring privacy through differential privacy-based text embeddings and anonymized representations.
- Recurrent neural networks (RNN): RNNs can model sequential data (e.g., time-series data, chat logs) while employing differential privacy to prevent the exposure of private patterns.
- Support vector machines (SVM): SVMs can classify sensitive data (e.g., medical conditions, financial fraud) while ensuring privacy by adding differential privacy noise to decision boundaries.
- Artificial neural networks (ANN): ANN-based models can process private datasets with DP-SGD to protect individual data points.
- Logistic regression: It can be used for private classification tasks (e.g., predicting user behavior) by incorporating differentially private optimization methods.
- Naïve Bayes: It can be used for privacy-preserving probabilistic classification by adding Laplace noise to probability distributions.
- K-nearest neighbors (KNN): It can classify data without explicitly storing individual records by using differential privacy mechanisms to prevent sensitive data from being queried.
- Deep learning: Various deep learning architectures (e.g., CNNs, LSTMs) can be trained with DP-SGD to ensure privacy during model training.

3 Methodology

According to the authors in [32], a well-structured review of literature should adhere to a coherent framework. Literature evaluations elucidate the progression of knowledge debates and offer valuable insights into authors' viewpoints [33]. A structured review, in particular, can provide targeted inquiries into knowledge organization and the historical development of a subject, enabling researchers to devise innovative approaches. Therefore, this study aligns with the guidelines outlined by [32] to address RQ1 and RQ2.

To gather information on the BC system's security, we conducted a comprehensive search across prominent publishers'databases, including Elsevier, Springer, Emerald, IEEE, MDPI, Taylor, and Google Scholar. Employing a search strategy using "AND" and "OR" logical operators, we aimed to identify relevant research articles about differential privacy and AI. Specifically, our search query

Table 1 Machine learning techniques used for differential privacy

Technique	Strengths	Weaknesses
NLP	Effective for text data	High-dimensional noise affects accuracy
RNN	Captures temporal dependencies	Gradient leakage risk in differential privacy training
SVM	Works well with small datasets	Poor scalability with differential privacy noise
ANN	Learns complex patterns	Noise in DP-SGD reduces accuracy
Logistic regression	Simple, efficient	Limited for complex relationships
Naïve Bayes	Works with small datasets	Assumes feature independence
KNN	Effective for instance-based learning	Computationally expensive with differential privacy
Deep learning	Handles complex data	Requires large datasets, differential privacy instability

"(artificial intelligence OR AI OR deep OR machine OR learning) AND (differential OR privacy)" was adopted, as detailed in [34]. In addition to the literature review, we delved into industrial websites, practitioner blogs, and professional reports to enrich our data collection [33]. Given the paper's concentration on ongoing and upcoming initiatives and projects related to the usage of differential privacy with AI, areas actively from governmental bodies and agencies, a portion of our insights are derived from industrial and professional websites.

While this paper does not follow a systematic literature review approach, our objective was to gather extensive information on the deployment of differential privacy in AI applications. However, we applied specific inclusion criteria to guide our search, limiting our sources to English-language publications that explicitly discuss AI or differential privacy. Our comprehensive search yielded a total of 48 resources, including 19 websites, professional blogs, or online links, five conference proceedings, and 24 journal articles. A summary of these findings is presented in Table 2.

Table 2 Summary of articles included in this paper

Article	Articles source	Total
[3, 7, 8, 10, 13, 15–17, 19, 21, 22, 26, 35–49]	Journal	24
[6, 50–53]	Conference proceedings	5
[11, 12, 14, 18, 20, 27, 28, 54–65]	Professional websites/blogs	19

4 Potentials of differential privacy and artificial intelligence combination

The combination of AI and differential privacy is summarized in Fig. 2 and discussed in the following subsections. These potentials were categorized into four major themes: enhancing privacy, responsible AI development, improving data sharing, and mitigating AI's bias concerns.

4.1 Enhancing privacy

Differential privacy enables AI models to be developed and used on sensitive data while mathematically protecting individual privacy. This is critical in industries such as healthcare and banking, where data security is vital. This

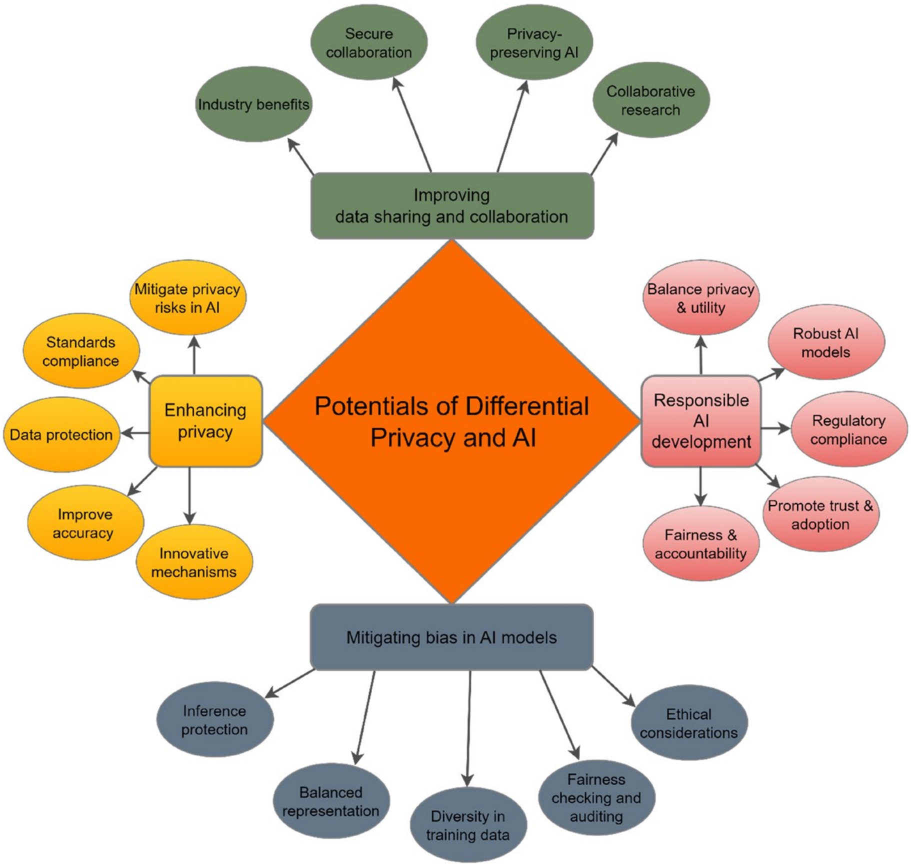

Fig. 2 Potentials of combining AI and differential privacy

theme has been reported in 13 references (27%) from the selected articles. Table 3 illustrates how differential privacy techniques can improve AI privacy.

By incorporating controlled noise into data or query results, differential privacy prevents the identification of specific individuals, even within large datasets, ensuring that individual data points remain confidential [14]. This added noise masks the details of any single individual's data point while allowing the model to learn general trends and patterns [42, 49]. The noise level can be adjusted to balance the desired privacy level and acceptable accuracy trade-off [64]. Differential privacy protects against inference attacks by ensuring that outputs do not reveal individual data points, thereby strengthening the security of AI systems [41]. It can also be integrated with federated learning, where AI models are trained across decentralized devices while keeping data localized, ensuring that even if local data is compromised, the overall privacy of individuals is maintained [19]. Moreover, this added noise allows AI models to learn patterns without compromising individual patient privacy [11, 49].

Many sectors are subject to rigorous data privacy requirements, including GDPR, HIPAA, and CCPA. Differential privacy enables enterprises to comply with these rules by providing measurable privacy assurances and ensuring that AI models manage data per legal norms. The authors in [58] emphasized the significance of reviewing and standardizing differential privacy approaches for reliable AI development [58]. Furthermore, the authors in [16] stressed the need for standardization and recommended practices for implementing differential privacy in real-world systems. It examines many issues and potential solutions to ensure the efficacy and dependability of differential privacy strategies [16].

Several studies have explored how differential privacy can be applied to NLP models, which often deal with sensitive information like text data [19]. The authors in [41, 50], and [46] explored specific differential privacy mechanisms that can be used to achieve enhanced privacy protection in AI applications. They studied the Lipschitz property-based differential privacy classification, the RNN-based differential privacy, and the privacy-aware trajectory generation model with differential privacy (differential privacy-TrajGAN), respectively. The authors in [19] highlighted the potential of differential privacy to not only protect privacy but also improve the overall utility (accuracy) of AI models in some cases. This can happen because adding noise can sometimes reduce the impact of biases present in the data, leading to more accurate models [11].

Implementing differential privacy in AI models is crucial for ensuring privacy when working with sensitive data, but several challenges arise in real-world applications. Differential privacy's core mechanism—adding noise to the data—can preserve privacy, but it may also reduce the model's accuracy if not carefully tuned [15]. The feasibility of enhancing privacy with differential privacy depends on careful tuning of the privacy-accuracy trade-off and computational resource allocation [64]. Its adoption in privacy-sensitive industries is growing, but the practical implementation still faces challenges.

4.2 Responsible AI development

Differential privacy plays a crucial role in promoting responsible AI development by mitigating the privacy risks associated with AI models. By addressing these risks, differential privacy fosters trust in AI systems and encourages their wider adoption, thereby promoting ethical and responsible AI development. Table 2 summarizes how differential privacy promotes responsible AI. This theme has been reported in 10 references (21%) from the selected articles. Table 4 summarizes how differential privacy techniques may enhance responsible AI models.

Differential privacy balances data usefulness and privacy by adjusting noise levels. This ensures that data remains relevant for training AI models while preserving individual privacy [20]. AI systems require stability for optimal performance [12]. Furthermore, differential

Table 3 Enhancing privacy

Potential	Recommendation	Study
Data protection	• Differential privacy prevents the identification of specific individuals	[14, 42, 49]
Mitigating privacy risks in AI	• Differential privacy defenses against inference attacks • Differential privacy allows AI models to learn patterns without compromising individual patient privacy	[11, 19, 41, 49]
Standards compliance	• Standardizing differential privacy techniques is essential for trustworthy AI development	[16, 58]
Improve the overall accuracy	• Differential privacy reduces the influence of biases in the data which results in more accurate AI models	[19, 64]
Innovative mechanisms	• Lipschitz property-based differential privacy classification	[50]
	• Differential privacy P-TrajGAN model	[46]
	• RNN-based differential privacy	[41]

Table 4 Enhancing responsible AI development

Potential	Recommendation	Study
Balancing data privacy and utility	• Ensuring that the data remains useful for training AI models while still protecting individual privacy • Minimizing the model's overfitting • Adjusting the amount of noise added	[12, 20, 49]
Regulatory compliance	• Following NIST standards may result in more fairness, accountability, and transparency	[28, 54, 66]
Robust AI model	• Regularizing AI models, which leads to improving their robustness • Commitment to responsible data handling and AI deployment	[15, 65]
Promoting trust and adoption	• Differential privacy promotes trust by minimizing the privacy concerns associated with AI models	[54, 61]
Enhancing fairness and accountability	• Defining probabilistic mappings from individuals to intermediate representations	[53]

privacy can reduce overfitting by reducing reliance on individual data points, leading to better generalization of previously unknown data [49]. Another benefit of differential privacy is the ability to adjust the amount of noise added to balance the trade-off between accuracy and privacy [55]. On the other hand, differential privacy encourages trust by demonstrating a reduction in privacy hazards associated with AI models [61]. Consumers who trust the safety of their data are more likely to share it with AI systems, perhaps leading to increased adoption of AI in numerous fields [54].

Regulations like the GDPR and the CCPA are compelling organizations to reduce gathered private data, prompting the increased implementation of differential privacy approaches [28]. Governments and regulatory authorities must monitor the development and deployment of AI technology by creating legislation, standards, and oversight bodies to guarantee responsible and ethical usage of AI while preserving individual privacy rights [66]. Regulations need to protect privacy and foster trust, ensuring transparency, accountability, and the individual's right to their data [54]. In Europe, the GDPR safeguards personal data from the threats of AI technology, while the FTC keeps corporations accountable for collecting, utilizing, and safeguarding their customers'data, punishing those that misrepresent data collection procedures [54].

Differential privacy includes privacy-preserving measures in data gathering and model training, allowing enterprises to demonstrate their commitment to responsible data management and AI implementation. It assures compliance with data protection standards and aids in the creation of an agreement on optimum privacy protection methods in AI development [15]. Differential privacy gives a mathematically precise definition of privacy, assuring that the outcomes of a statistical study are "essentially indistinguishable" whether or not a single human is included in the dataset. This rigor helps preserve individual privacy even when data is harvested for analysis [65]. Moreover, differential privacy systems attempt to accomplish both group and individual fairness by creating probabilistic mappings from persons to intermediate representations. This assures that the proportion of members in a protected group obtaining positive categorization is the same as the population's proportion and that comparable people are treated equally. However, the lack of clear measures to evaluate individuals makes reaching this goal impossible [53].

Differential privacy plays a critical role in ensuring responsible AI development by mitigating privacy risks and building trust in AI systems. However, while differential privacy plays a vital role in fostering responsible AI, the feasibility of enhancing responsible AI implementation requires clear regulations, effective communication to build trust, and a careful balancing of privacy with interpretability and model performance [12, 21].

4.3 Improving data sharing and collaboration

The combination of AI and differential privacy improves data sharing across businesses by ensuring robust privacy while allowing for the extraction of important insights. This integration promotes cooperation, creativity, and efficiency while ensuring confidence and adherence to privacy standards. As a consequence, companies may use data-driven decision-making to enhance services, streamline operations, and provide tailored experiences while maintaining individual privacy. Differential privacy promotes secure collaboration on AI projects by allowing researchers to share data without jeopardizing individual privacy. This can boost innovation in a variety of disciplines. This theme has been reported in 7 references (14.5%) from the selected articles. Table 5 illustrates how differential privacy techniques may enhance data security.

Differential privacy enables researchers and businesses to safely exchange sensitive data for collaborative AI projects, boosting innovation while protecting individual privacy [15]. Collaborative learning approaches

Table 5 Improving data sharing

Potential	Recommendation	Study
Enabling secure collaboration	• Differential privacy can facilitate secure collaboration on AI projects • Mathematically guaranteeing privacy protection	[15, 64]
Privacy-preserving AI	• Enabling the development and deployment of various models • Preventing the reconstruction and inference of sensitive data	[10, 52]
Broader industry benefits	• Healthcare, finance, government, retail, etc • Faster innovation and development	[48, 57]
Collaborative research	• Data and identities remain protected throughout these collaborations • Ethical data management	[65]

with differential privacy allow secure cooperation on AI projects while mathematically ensuring privacy protection [64]. This facilitates faster innovation in fields like healthcare and finance, where data collaboration is crucial for progress. Differential privacy quantifies privacy assurances without assuming specific attack models, and it provides a uniform assessment criterion for privacy-enhancing solutions, ensuring privacy protection stays effective even as attackers' computing resources increase [15]. Differential privacy provides a comprehensive approach for safeguarding individual privacy while maintaining dataset usability, which is critical for database owners. This allows the publishing and management of datasets for societal advantages while adhering to ethical privacy standards [65].

Differential privacy enhances machine learning by incorporating privacy-preserving mechanisms. It enables the development and deployment of various models, including support vector machines, artificial neural networks, logistic regression, naive Bayes, and k-nearest neighbors, with high prediction accuracy while ensuring privacy [52]. Training deep learning models in a differentially private manner prevents the reconstruction and inference of sensitive information, enhancing data security [10].

Banks and fintech firms may leverage differentially private data to deliver personalized financial goods and services, such as tailored investment advice and customized lending solutions while protecting consumer privacy [48]. Retailers may share anonymized transaction data with manufacturers and marketers to learn about customer behavior, forecast trends, improve inventory, and customize marketing efforts. Governments can make anonymized census and public service data available to researchers and policymakers. AI models may use this data to influence governmental policy, improve social services, and address societal concerns while protecting individual privacy. Educational institutions can exchange student performance information while retaining anonymity. AI models may use this data to discover learning trends, create individualized learning strategies, and enhance educational outcomes [57]. In general, differential privacy can significantly improve data sharing and collaboration between organizations while protecting privacy. However, the feasibility of this collaboration requires addressing sector-specific privacy needs, ensuring data utility, and overcoming technical challenges in distributed environments [28, 35].

4.4 Mitigating bias in AI models

Differential privacy, which protects individual data, can help reduce biases in AI algorithms. It prevents outliers and unique data points from influencing models, resulting in more fair and impartial results. Differential privacy can help decrease bias in AI models trained on sensitive data by providing noise to disguise any biases in the data. This theme has been reported in 6 references (12.5%) from the selected references. Table 6 summarizes how differential privacy techniques may mitigate the bias in AI models.

Differential privacy guarantees that sensitive properties are not inferred from the dataset, even if an attacker has access to external information [15]. This decreases the possibility of models unintentionally learning and perpetuating biases seen in the training data. The noise supplied to the data during differential privacy prevents the model from overfitting to individual data points that may include inherent biases. This results in models that are more generalizable and less susceptible to bias from outliers [57]. Furthermore, differential privacy guarantees that the contributions of individual data points, particularly those from minority groups, are equal. This prevents models from being dominated by the majority class and promotes equitable representation of all groups in the dataset. In datasets with noisy labels, minority groups may be misrepresented owing to inaccurate labeling [20]. Differential privacy helps to mitigate the influence of these noisy labels, resulting in more accurate and unbiased model training. It facilitates the inclusion of varied and sensitive material in training datasets while

Table 6 Mitigating bias in AI models

Potential	Recommendation	Study
Inference protection	• Preventing attribute inference • Reducing overfitting and labeling	[15]
Balanced representation	• Preventing majority class domination • Inclusion of diverse and sensitive data • Reducing the impact of noisy labels	[20]
Enhancing diversity in training data	• Safe inclusion of diverse data • Encouraging data contributions	[20]
Fairness checking and auditing	• Transparent evaluation • Integration with fairness constraints and dynamic adjustments	[48]
Ethical considerations	• Differential privacy ensures that AI models are deployed ethically • Ensuring same inferences about any individual's private data	[15, 57]

maintaining privacy [20]. This inclusion helps to create more representative and impartial models. When people feel their privacy is respected, they are more willing to give their data, resulting in larger and more varied datasets, which are critical for training impartial AI models [48].

Differential privacy enables companies to exchange anonymized data for bias audits and fairness checks while protecting individual privacy. External auditors can assess the fairness of AI models and give input for improvement [20]. Organizations may be more open about how they gather and manage data by releasing differentially private datasets. This transparency aids in detecting and correcting biases in data and models. Differential privacy can be used with fairness-aware learning algorithms that are specifically designed to reduce prejudice. These integrated techniques ensure that privacy and fairness are maximized simultaneously throughout model training. Differential privacy enables dynamic changes to the training process to overcome emergent biases. This flexibility contributes to ensuring fairness throughout the model's existence [48].

Promoting ethical concerns in data processing and model training is a significant benefit of including differential privacy in AI development [15]. Integrating differential privacy into AI development complies with ethical data-use guidelines, guaranteeing that AI systems respect user privacy and function transparently [15]. This encourages the creation of AI systems that are socially and morally sound. Differential privacy is an important assurance that persons who see the results of a differentially private analysis will typically make the same inferences about any individual's private information—even if that person's personal information isn't included in the analysis's input [57].

Differential privacy can help mitigate bias in AI models by ensuring that no single data point disproportionately affects the model's learning process. However, its feasibility in addressing bias in AI models depends on various factors including the quality of the training data and the need for additional fairness-aware approaches [15]. It can be a tool for reducing bias, but it is not a comprehensive solution on its own [5].

5 Challenges of differential privacy and artificial intelligence combination

Integrating differential privacy with AI offers significant potential for enhancing privacy protection while leveraging the power of AI for data analysis. However, AI and its combination with differential privacy are still in their early phases, making them vulnerable to several challenges. This section discusses the concerns and potential solutions for this combination (RQ2). The challenges of this combination are summarized in Table 7 and discussed in the following subsections.

5.1 Accuracy vs. privacy trade-off

Accuracy refers to the discrepancy between the output of a mechanism and the real value that it is attempting to approximate [5]. The trade-off between accuracy and privacy when integrating differential privacy with AI models necessitates novel methodologies and ongoing research to ensure that both privacy and utility are maximized without jeopardizing AI model performance. Adding noise to data for privacy can degrade AI model accuracy. This trade-off is a key difficulty in obtaining high privacy while also providing great value. Balancing privacy and accuracy are crucial since increasing privacy frequently affects data usefulness and model accuracy [47]. In the following paragraphs, we discuss the recommendations to achieve this balance.

- Balancing privacy and utility: Differential privacy involves adding noise to data in order to safeguard individuals' privacy. However, this noise might limit the value of the data, making it less precise and valu-

Table 7 Potentials of integrating AI with differential privacy

Challenge	Recommended mitigation	Study
Accuracy vs. privacy trade-off	• Finding the right balance between privacy and accuracy • Balancing between large and small datasets • Deploying adaptive privacy mechanisms • Optimizing privacy parameters • Deploying hybrid models and advanced techniques • Government and institutional approaches	[15, 18, 37, 45, 47, 64]
Computational complexity	• Addressing computational overhead • Leveraging advanced techniques and tools • Utilizing libraries and frameworks • Training and collaboration • Optimization for large datasets • Accessibility issues	[19, 21, 37]
Evolving regulatory landscape	• Complexity of compliance • Risk management • Communication challenges • Compatibility with existing systems • Developing flexible frameworks • Data governance policies • Role of standards and guidelines • Effective regulation and oversight	[12, 15, 21, 37]
Limited expertise	• Shortage of skilled professionals • Collaborations and resources • Technical challenges • Investing in training and development • Sharing knowledge and best practices • Educating AI developers • Enhancing transparency and explainability	[15, 28, 37, 40]
Data utility and usability	• Balancing privacy and utility • Data distortion risk • Advanced noise-reduction techniques • Improved algorithms • Differentially private synthetic data	[20, 28, 35, 47, 64]
Scalability issue	• Real-time transmission • Efficient algorithms and data structures • Parallel and distributed computing techniques • Dynamic privacy budget allocation	[10, 18, 19, 44, 60]
Bias issue	• Thorough monitoring and assessments • Fairness-aware algorithms • Diverse and representative datasets • Ethical considerations	[5, 15, 18, 20, 38]

able for certain sorts of analyses. This trade-off can be difficult to handle and requires a careful balance of both privacy and function [18]. There is a trade-off between convergence performance and privacy protection levels, which means that greater convergence performance leads to a lower protection level [45].

- Balancing between large and small datasets: The difficulty of combining privacy and accuracy is especially important in sophisticated activities such as picture recognition. Adding noise to preserve privacy may reduce the accuracy of AI models. Research into more computationally efficient differential privacy algorithms is critical for AI adoption on a broad scale, with the goal of improving efficiency while maintaining privacy guarantees [47]. Furthermore, the inaccuracy caused by differential privacy after adding noise to the dataset may be insignificant for big datasets but not for small ones. The reorganized data after using differential privacy algorithms might impede analysts from discovering significant insights from the data presented [64].
- Adaptive privacy mechanisms: To solve the accuracy vs. privacy dilemma, it is important to use adaptive privacy techniques that modify the amount of noise introduced based on data sensitivity and desired privacy levels [45]. This technique enables fine-tuned control of the trade-off.
- Optimizing privacy parameters: It is critical to determine ideal privacy settings that ensure a reasonable amount of privacy while maintaining the effectiveness of AI algorithms. To keep AI models accurate and successful, privacy protection must be balanced with data usefulness [15].

- Hybrid models and advanced techniques: Hybrid models that use both differentially private and non-private data can help maintain better accuracy while preserving a baseline degree of privacy accuracy [18]. Incorporating sophisticated machine learning approaches, such as transfer learning or semi-supervised learning, can also help enhance model accuracy in the presence of noise. Advanced noise reduction techniques and refined algorithms can reduce the influence on accuracy [18]. Adaptive noise mechanisms that modify noise levels according to data sensitivity help balance privacy and accuracy.
- Government and institutional approaches: Synthetic data, both old synthetic data and novel synthetic data made utilizing differentially private approaches, are frequently insufficient to develop sophisticated statistical models [18]. To address this problem, the United States government established Federal Statistical Research Data Centers, which allow adequately verified researchers to access secret data and share results following acceptable disclosure approval processes at federal statistical agencies. This technique is frequently lengthy, arduous, and underused [37].

5.2 Computational complexity

Integrating differential privacy into AI models can be technically difficult, creating a challenge, particularly for firms with little knowledge in both domains. Implementing differential privacy methods frequently takes large computer resources, slowing down the training and inference procedures [20]. The complexity of differential privacy approaches involves significant computing costs, which impact the efficiency and scalability of AI systems [15]. In the following paragraphs, we discuss the recommendations to mitigate the computational complexity.

- Addressing computational overhead: To control the computational cost associated with differential privacy, it is critical to create and use optimal algorithms built expressly for this purpose [21]. These methods should seek to decrease computational complexity while maintaining strong privacy guarantees.
- Leveraging advanced techniques and tools: Using hardware acceleration, such as GPUs or specialized CPUs for machine learning applications, may drastically cut calculation time. Parallel processing and distributed computing solutions can help manage the workload more efficiently [37]. Using approximation approaches or subsampling techniques can give faster computations while preserving an appropriate level of confidentiality and accuracy.
- Utilizing libraries and frameworks: To reduce implementation complexity, libraries and frameworks with built-in support for differential privacy are recommended. These technologies frequently abstract most of the complexity and provide standardized mechanisms for developing privacy-preserving strategies [21].
- Training and collaboration: Investing in training and education for data scientists and engineers is critical to developing competence in differential privacy. Collaborating with academic institutions or industry specialists can give further help and direction [19]. By simplifying the architecture and focusing on modular implementations, the incorporation of differential privacy may become more manageable and error-free.
- Optimization for large datasets: Implementing differential privacy techniques may be computationally costly, particularly with huge datasets. To address computational problems, these methods must be optimized for efficiency. The research underlines the necessity to address the computational complexity of several differential privacy strategies and the issues associated with using them on large datasets [19].
- Accessibility issues: Many data practitioners and users struggle to execute differentially private approaches due to low computing resources. These constraints impede accessibility for the common data consumer, who may not have the necessary computing power to execute the techniques or the background to hand-code them [37].

5.3 Evolving regulatory landscape

Managing the dynamic regulatory landscape when integrating differential privacy with AI models necessitates a multidimensional strategy that includes adaptable frameworks, strong governance principles, effective communication, and stakeholder participation. This strategy guarantees compliance, encourages ethical AI development, and safeguards individual rights and liberties. In the following paragraphs, we discuss the recommendations to mitigate the regulation issue.

- The complexity of compliance: Ensuring compliance with various and growing privacy rules across several areas may be difficult. Organizations must be educated on the most recent privacy laws and regulations in various jurisdictions to overcome regulatory and compliance problems [21].
- Risk management: Without adequate regulation, the expanding use of AI technology risks further eroding privacy and civil rights, as well as worsening soci-

etal imbalances and prejudices. Establishing a legal framework for AI can help guarantee that this powerful technology is used for the greater good while respecting individuals' rights and liberties [12].

- Communication challenges: One of the most difficult issues in implementing differentiated privacy is conveying to legislators and the general public the underlying mathematics of privacy based on what we have learned over the last two decades. Such conversations are problematic because the security of private information operates differently than our intuition has been educated to believe [37].
- Compatibility with existing systems: Ensuring interoperability with existing AI systems and workflows without severe interruptions is critical for preserving operational efficiency while including differentiated privacy protections [15].
- Developing flexible frameworks: Creating a flexible privacy architecture that can rapidly adjust to changing legislative needs is critical. Engaging legal experts and compliance officials during the design and implementation phases may help guarantee that the system complies with all applicable legal requirements. Regular audits and reviews of the privacy framework can help detect and handle compliance concerns more proactively [15].
- Data governance policies: Implementing strong data governance rules and processes will ensure long-term compliance with privacy requirements. Privacy requirements are continually changing, demanding flexible and adaptive differential privacy solutions to ensure compliance. Establishing industry-wide standards and best practices for combining differential privacy with AI ensures consistency and efficacy [15].
- Role of standards and guidelines: Organizations like NIST may help with compliance by offering rules and frameworks. Clear rules and governance frameworks describing the ethical use of differential privacy in artificial intelligence, along with frequent audits and assessments, can help ensure adherence to privacy standards and best practices [61].
- Effective regulation and oversight: To guarantee that AI technology is created and utilized in a way that protects individual rights and liberties, it must be subject to effective regulation and control. This encompasses not just the collecting and use of data by AI systems but also the design and development of these systems to guarantee they are visible, explainable, and objective [12]. Effective regulation of AI technology would need collaboration among governments, industry, and civil society to develop clear rules and guidelines for its ethical usage. Continuous monitoring and enforcement are required to guarantee that these criteria are met [12].

5.4 Limited expertise

Addressing the lack of competence in merging differential privacy with AI models requires investing in training and development, encouraging collaboration, raising awareness, and overcoming technological hurdles. These procedures will enable the effective implementation and widespread use of privacy-preserving strategies in AI systems. In the following paragraphs, we discuss the recommendations to mitigate the limited expertise issue.

- Shortage of skilled professionals: There is frequently a scarcity of individuals with the essential knowledge and abilities to successfully deploy differential privacy approaches in AI systems [28]. This lack may impede the adoption and appropriate implementation of privacy-preserving technologies.
- Collaborations and resources: Collaboration with academic institutions and industry specialists to create training materials and undertake collaborative research can also be advantageous. Encouraging a culture of continual learning and giving access to tools such as research papers, online courses, and professional groups may help staff keep on top of the newest innovations [40].
- Technical challenges: The shortage of practitioners who are both mathematicians and excellent builders of production systems is perhaps the most significant impediment to the implementation of differential privacy. Although there are a rising number of production-ready differential privacy libraries, such as Google's Privacy on Beam, the process of developing, creating, deploying, and maintaining a functioning differential privacy system needs significantly more than a verified differentially private algorithms library [40]. For example, the practitioner must determine which algorithms to utilize and where to place them in the statistical pipeline. Frequently, the practitioner must revise statistical computations in order to use these data more efficiently [37].
- Investing in training and development: Raising awareness of differential privacy and its benefits among AI developers and stakeholders is critical to its widespread adoption. Organizations may solve the dilemma of limited knowledge by investing in comprehensive training and development programs to upskill their current staff [15]. Offering seminars, courses, and certifications on differential privacy and its applications in AI can assist in closing the knowledge gap. Hiring professionals in differential privacy

and AI, or consulting with specialized businesses, can also provide the essential experience to lead and assist the implementation [15].

- Sharing knowledge and best practices: Fostering collaboration between differential privacy experts and AI developers can hasten innovation and practical application. This collaboration can help to close the knowledge gap and create more user-friendly solutions. Encouraging collaboration among academics, businesses, and regulatory organizations to share knowledge and best practices is also critical [28]. Creating open-source tools and frameworks that ease the integration of differential privacy into AI systems, as well as producing thorough documentation and tutorials to assist the adoption of privacy-preserving algorithms, are essential steps in addressing the issue of limited knowledge [37].
- Educating AI developers: Educating AI developers and practitioners on the implementation and consequences of differential privacy is necessary to enable appropriate AI integration. Addressing the complexity of assessing the efficiency of differential privacy in AI systems, particularly in real-world scenarios with changing data environments [15].
- Enhancing transparency and explainability: Improving transparency and explainability in AI models with differential privacy is critical to promoting accountability and trustworthiness in data processing and decision-making processes [15].

5.5 Data utility and usability

Utility measures the usefulness of a dataset or statistic for a certain purpose [5]. Addressing the difficulties of data usefulness and usability in differential privacy when integrated with AI models entails adopting sophisticated noise-reduction techniques, federated learning, producing differentially private synthetic data, and building better algorithms. These solutions strive to reconcile privacy with utility, guaranteeing that differentially private material is nonetheless relevant for AI model training. In the following paragraphs, we discuss the recommendations to mitigate the data utility issue.

- Balancing privacy and utility: Keeping differentially private data valuable for AI model training while respecting privacy is difficult. Achieving the appropriate balance between privacy and utility involves knowledge and computationally intensive hyperparameter adjustment [28].
- Data distortion risk: While differentially private procedures are used to safeguard privacy, the approaches must be successful in realistic, real-world settings. Continuous improvements and breakthroughs in differential privacy strategies are required to achieve an optimal balance between privacy protection and data value. While the purpose is to avoid distortion, the noise addition might occasionally undermine the data's value, especially when privacy parameters are not properly controlled [20].
- Advanced noise-reduction techniques: Advanced noise-reduction techniques can be used to address challenges related to data usefulness and usability. Adaptive noise methods that alter the quantity of noise based on data sensitivity and the AI model's unique requirements assist in maintaining a balance between privacy and data usefulness [19]. Additionally, post-processing techniques can modify noisy outputs, preserving the data's key properties while reducing further noise. Perturbation approaches that introduce little noise to data can also keep it usable for AI applications while protecting privacy. Adding noise selectively to the most sensitive characteristics of the data can lessen the overall impact on the data value [20].
- Improved algorithms: Improved algorithms and various privacy metric measures are presented to cope with differential privacy for correlated data, falling under dependent differential privacy [64]. Federated learning is another efficient way that allows AI models to be trained locally on user devices without the need to exchange raw data, increasing data privacy while maintaining utility. By aggregating models from many sources in a privacy-preserving way, the global model gains access to varied data while maintaining individual privacy. Dimensionality reduction can focus on the most informative aspects, protecting data utility and privacy [47].
- Differentially private synthetic data: Experiments have demonstrated that differential privacy does not perform well with databases built on tuple correlation, which represents relationships between distinct tables in the database. Differential privacy also modifies records that do not include numerical or statistical trends, i.e., it can only introduce noise to records with categorical data [47]. Privacy-preserving data transformations, such as differentially private synthetic data creation, provide new datasets that maintain the statistical features of the original data, allowing for successful model training without disclosing sensitive information [35].

5.6 Scalability issues

Scaling differential privacy to accommodate vast datasets and complicated AI models may be difficult.

Maintaining privacy while processing large volumes of data is complicated and computationally demanding. Scalability in differential privacy requires efficient algorithms, parallel and distributed computing, cloud platforms, incremental learning, and batch-processing approaches. Dynamic privacy budget allocation and secure privacy budget reuse improve scalability even more. Real-time transmission in smart cities and increasing IIoT applications require special measures to ensure privacy protection keeps up with technical improvements. In the following paragraphs, we discuss the recommendations to mitigate the scalability issue.

- Real-time transmission: Designing differential privacy algorithms for real-time transmission in IoT-based smart cities is a substantial problem. To ensure that data may be sent and exchanged between nodes within a specific latency range while retaining privacy, efficient and resilient algorithm designs are required [19].
- Efficient algorithms and data structures: Scalability issues can be solved by implementing efficient algorithms and data structures. Creating optimal differential privacy algorithms that are computationally efficient and capable of managing vast amounts of data is critical. These techniques should reduce the computational expense while maintaining privacy [19]. Sparse data structures can minimize memory and processing requirements, making differential privacy approaches more scalable for large datasets. Furthermore, batch processing of data can help manage memory and computing resources more effectively. Processing data in batches provides better control over resource utilization and can increase the scalability of differential privacy methods [10].
- Parallel and distributed computing techniques: Using parallel and distributed computing techniques can dramatically improve scalability. Implementing distributed privacy methods enables data to be handled in parallel across numerous nodes, decreasing the computing strain on any single node [60]. Cloud computing systems provide scalable resources for processing massive datasets and executing complicated AI models while maintaining differentiated privacy. Cloud platforms provide the infrastructure required to efficiently conduct large-scale calculations [36]. Incremental learning approaches, in which the AI model is updated progressively with fresh data batches, can lessen the need to reprocess the whole dataset while increasing scalability. This strategy helps to manage the computational burden more effectively [44].
- Dynamic privacy budget allocation: Dynamic privacy budget allocation solutions distribute the privacy budget based on the significance and amount of the data being processed. This technique facilitates effective management of the total privacy budget across huge datasets. Techniques for properly recycling sections of the privacy budget across multiple activities or phases of AI model training can maximize the use of available resources, further addressing scalability difficulties [18].

5.7 Bias and fairness concerns

Differential privacy might unintentionally add or worsen bias in AI algorithms, potentially resulting in unjust conclusions. Addressing bias and fairness problems in differential privacy for AI models requires a diverse strategy [38]. This involves in-depth evaluations, fairness-aware algorithms, diversified datasets, frequent monitoring, stakeholder participation, and a heavy emphasis on ethical concerns [39]. Overcoming the problems of standardization will be critical for the broad and fair application of differentiated privacy in AI systems. In the following paragraphs, we discuss the recommendations to mitigate the bias issue.

- Thorough monitoring and assessments: Regularly monitor and update the models to address developing bias concerns [20]. Continuous review promotes justice and equity in AI systems. Conduct rigorous reviews of AI models before and after implementing differential privacy approaches. This helps to identify and address any bias that may be established during the process [15].
- Fairness-aware algorithms: Creating and implementing fairness-aware differential privacy algorithms may explicitly account for and reduce prejudice. These algorithms ensure that privacy-preserving measures don't have disproportionate effects on certain demographics [38].
- Diverse and representative datasets: Using varied and representative datasets to train AI models may lower the risk of bias by ensuring that the training data reflects a broad range of the population [5].
- Ethical considerations: Emphasize ethical issues while designing and deploying AI systems that employ differential privacy. This entails involving stakeholders, including end users, in conversations about privacy expectations and the ethical implications of data usage. Ethical issues should be incorporated into the development process to guarantee that AI systems are both technically sound and socially responsible. Regularly assessing and upgrading the

models to integrate ethical rules and fairness principles will help preserve the integrity and fairness of the AI systems throughout time [18].

6 Discussion

This study addressed two research questions, namely the potentials and challenges of combining differential privacy with AI. This section addresses the third research question related to the future direction of this combination by focusing on the most common implications found in this paper. This section also addresses the research limitations.

6.1 Implications and future directions

Both theoretical and practical implications underline the significance of striking a balance between privacy and value, encouraging cooperation, and resolving prejudice and fairness issues. Future research and practical efforts should focus on building efficient algorithms, standardizing methods, and encouraging ethical and responsible AI development.

6.1.1 Theoretical implications and directions

Several theoretical implications were identified in this research. Researchers may use differential privacy with AI to use AI's capability for data analysis while protecting individual privacy. This technique has great potential for different industries that rely on sensitive data, supporting responsible and ethical AI development. However, obtaining the ideal balance of privacy and utility (AI model accuracy) remains a huge theoretical challenge. Finding the right level of noise addition is an ongoing field of research, underscoring the necessity for continuing investigation into the balance between privacy and utility.

Moreover, differential privacy facilitates secure data exchange between researchers and organizations. By ensuring that shared data cannot be traced back to people, it encourages collaborative study and innovation across disciplines. This facilitates an open data program, allowing public and commercial institutions to share anonymized datasets with the academic community while maintaining anonymity. However, most of the present research on differential privacy is mostly theoretical. Some proposed algorithms can only be utilized with data that meets specified requirements that are seldom seen in real-world circumstances, making practical implementations difficult. Future research should focus on constructing more computationally efficient and resilient differential privacy algorithms capable of handling varied data types and real-world settings.

Furthermore, differential privacy may unintentionally induce or worsen bias in AI models. The theoretical implications include creating fairness-aware differential privacy algorithms that explicitly account for and minimize bias. This necessitates a thorough knowledge of how noise addition might affect certain groups, as well as the creation of mechanisms to ensure that privacy-preserving procedures do not disproportionately influence specific populations. The continued development of fairness-aware algorithms and strategies for monitoring and adjusting for bias in differentially private models is critical. To ensure societal responsibility and fairness, researchers should incorporate ethical issues into the development and deployment of AI systems, leveraging differential privacy.

Finally, the lack of standardized and agreed-upon best practices remains a theoretical obstacle. Establishing industry-wide standards and best practices for incorporating differential privacy with AI is critical to ensuring consistency and efficacy. This necessitates collaboration among scholars to establish generally applicable recommendations. Collaborative efforts to develop established principles and best practices for enabling differential privacy in AI are necessary.

6.1.2 Practical implications and directions

Several practical implications were also identified in this paper. First, differential privacy encourages secure data exchange across researchers and institutions, allowing joint study and innovation in several sectors [27]. This is especially useful in healthcare, where data is frequently sensitive and private. However, differential privacy is most successful in complicated use cases when combined with a comprehensive data security framework. Raising awareness and education on differential privacy among AI developers and stakeholders is critical to broader adoption. Offering training programs, workshops, and resources can help close the knowledge gap. Standardized frameworks and best practices for integrating differential privacy in real-world applications will encourage further adoption.

Moreover, the use of differential privacy in real-world applications, such as the 2020 US Census, illustrates its practical utility [65]. This establishes a precedent for future datasets, demonstrating how differential privacy may improve data privacy and statistical precision in censuses and other large-scale data-gathering tasks. AI models must be monitored and updated regularly to overcome growing biases and preserve fairness over time.

Furthermore, expanding differential privacy to accommodate vast datasets and complicated AI models can be difficult. Practical solutions rely on efficient algorithms, data structures, and distributed computing approaches. Implementing distributed privacy methods and leveraging cloud computing platforms provide scalable resources for processing massive datasets and running complicated AI models while maintaining differentiated privacy. Practical efforts should focus on building optimal differential privacy algorithms that are computationally efficient and capable of processing massive amounts of data.

Finally, to address bias and fairness problems in practice, AI models must be thoroughly evaluated before and after differential privacy approaches are applied. Engaging stakeholders, particularly affected communities, may give useful insights into ensuring that models are fair and equitable. Practical implementations should incorporate ethical norms and fairness concepts into the development process to guarantee that AI systems are socially accountable and transparent.

6.2 General recommendations

The integration of differential privacy and AI remains an evolving field, requiring further research and experimentation. In this section, we offer several recommendations for combining AI with differential privacy techniques. By following these recommendations, we hope that organizations can effectively integrate differential privacy with AI, ensuring robust data protection and maintaining the utility and fairness of AI models.

- Understand differential privacy: Differential privacy promises to protect individuals from additional harm caused by their data being stored in a database. Furthermore, differential privacy relates to queries rather than databases themselves. However, it does not guarantee that all secrets will stay confidential, especially against specific attacks such as differential attacks.
- Manage noise addition: More noise facilitates privacy while decreasing the accuracy and value of the AI model. To achieve the best mix of privacy and utility, epsilon and delta values may be fine-tuned through repeated iterations and validation.
- Handle privacy budget carefully: Injecting noise repeatedly across various searches might reduce output utility and deplete the privacy budget. Excessive noise and privacy budget exhaustion might impede future data utilization and undermine user requests.
- Seek expert consultation: The inclusion of data privacy professionals, legal advisers, and domain-specific experts may aid in properly applying privacy parameters.
- Implement regular audits and monitoring: Regular monitoring aids in adaptation to new security risks and assures the effectiveness of privacy protections.
- Establish comprehensive data governance: Consider data acquisition, storage, usage, and deletion, including ethical data use, privacy impact evaluations, and data breach processes.
- Practical steps for applying differential privacy: The following are good hints to consider: choosing an appropriate differentially private mechanism, such as the Laplace mechanism; calculating the privacy budget and determining function sensitivity; releasing the noisy version of the data; and applying the chosen mechanism to add noise while protecting individual privacy, which ensures it does not reveal specific individual information.
- Promote continuous learning and adaptation: This may be accomplished by addressing challenges and constraints as they arise, as well as keeping current on the latest research and best practices. This successfully protects data privacy while preserving utility.

6.3 Research limitations and future directions

This study acknowledges several limitations that could affect the comprehensiveness and applicability of its findings. First, the body of academic research specifically addressing the combination of differential privacy and AI remains relatively sparse. As a result, this study may not fully capture the breadth of theoretical insights or emerging trends at this intersection. The scarcity of peer-reviewed studies limits the ability to draw robust conclusions and underscores the need for further academic exploration. Moreover, while there is a growing pool of technical literature on differential privacy and AI individually, there is a notable gap in resources that comprehensively cover their integration. Many of the existing technical papers focus on isolated aspects or provide highly specialized information that may not be broadly applicable. This fragmentation can pose challenges in synthesizing a holistic view of best practices and common pitfalls when combining these technologies. The practical application of differential privacy in AI systems is still new, with few documented real-world implementations. This limitation restricts the study's ability to offer concrete, field-tested examples or case studies that demonstrate successful integration. Without substantial empirical evidence, the recommendations and insights provided may rely heavily on theoretical frameworks and hypothetical scenarios, which may not fully reflect the complexities encountered in practical settings.

7 Conclusions

As AI technology and its applications expand and become more generally utilized, it is essential to improve the privacy of the data they interact with. Adopting differential privacy notions is one promising strategy that has been advocated. However, there are limited studies on this combo. As a result, our study addressed this gap by identifying the potential, obstacles, and future directions for this combination. We investigated the literature available to address this gap; however, we observed that academic material was insufficient, so we extended our search to professional websites, blogs, and governmental connections. Our research has shown that combining differential privacy with AI has considerable promise for improving data privacy, responsible AI models, data sharing, and reducing bias and unfairness using AI models. Differential privacy allows for greater adoption of AI across multiple areas by resolving privacy concerns, guaranteeing that AI systems are both useful and ethically sound. This integration is crucial for the future of AI, as the demand for strong data-driven insights must be weighed against the need to preserve individual privacy. However, challenges were recognized and explored, including the accuracy-privacy trade-off, computational complexity, regulatory requirements, limited knowledge, scalability, data usefulness, and bias concerns.

Acknowledgements
Not applicable.

Authors' contributions
A.M. conceptualized and presented the idea. Y.A. developed the theoretical analysis. A.M. supervised the findings of this research. Y.A. took the lead in writing the manuscript with participation of A.M. Y.A. and A.M. reviewed the final version of the manuscript. All authors provided critical feedback and help shaping the research, results discussion and contributed to the final manuscript.

Funding
Open access funding provided by NTNU Norwegian University of Science and Technology (incl St. Olavs Hospital - Trondheim University Hospital)

Data availability
No datasets were generated or analysed during the current study.

Declarations

Competing interests
The authors declare no competing interests.

Received: 28 January 2025 Accepted: 21 May 2025
Published: 29 May 2025

References

1. G.S. Kumar, K. Premalatha, Securing private information by data perturbation using statistical transformation with three dimensional shearing. Appl. Soft Comput. **112**, 107819 (2021)
2. G.S. Kumar, K. Premalatha, STIF: Intuitionistic fuzzy Gaussian membership function with statistical transformation weight of evidence and information value for private information preservation. Distributed and Parallel Databases **41**, 233–266 (2023)
3. T. Wang, X. Zhang, J. Feng, X. Yang, A comprehensive survey on local differential privacy toward data statistics and analysis. Sensors **20**, 7030 (2020)
4. G.S. Kumar, K. Premalatha, G.U. Maheshwari, P.R. Kanna, No more privacy concern: a privacy-chain based homomorphic encryption scheme and statistical method for privacy preservation of user's private and sensitive data. Expert Syst. Appl. **234**, 121071 (2023)
5. J.P Near, Darais, D., Lefkovitz, N., Howarth, G. Guidelines for evaluating differential privacy guarantees. (2025). https://csrc.nist.gov/pubs/sp/800/226/final. Accessed 2 July 2025.
6. M. Ahmadzai, G. Nguyen, Federated learning with differential privacy on personal opinions: a privacy-preserving approach. Procedia Computer Science **225**, 543–552 (2023)
7. Y.I. Alzoubi, A. Mishra, A.E. Topcu, Research trends in deep learning and machine learning for cloud computing security. Artif. Intell. Rev. **57**, 132 (2024)
8. Z. Bu, H. Wang, Z. Dai, Q. Long, On the convergence and calibration of deep learning with differential privacy. Transactions on machine learning research **2023**, 1–36 (2023)
9. Y.I. Alzoubi, A. Mishra, A.E. Topcu, A.O. Cibikdiken, Generative artificial intelligence technology for systems engineering research: contribution and challenges. International Journal of Industrial Engineering and Management **15**, 169–179 (2024)
10. K. Pan, Y.-S. Ong, M. Gong, H. Li, A.K. Qin, Y. Gao, Differential privacy in deep learning: a literature survey. Neurocomputing **589**, 127663 (2024)
11. A. McFarland. What is differential privacy? https://www.unite.ai/what-is-differential-privacy/, accessed 1 July 2024. 2022.
12. D. MacRae. AI news: 80% of AI decision makers are worried about data privacy and security. https://www.artificialintelligence-news.com/2024/04/17/80-of-ai-decision-makers-are-worried-about-data-privacy-and-security/, accessed 1 July 2024. 2024.
13. Z. Huang, R. Hu, Y. Guo, E. Chan-Tin, Y. Gong, DP-ADMM: ADMM-based distributed learning with differential privacy. IEEE Trans. Inf. Forensics Secur. **15**, 1002–1012 (2019)
14. CLAN. Differential privacy in AI. https://clanx.ai/glossary/differential-privacy-in-ai, accessed 2 July 2024. 2024.
15. K. Kan, Seeking the ideal privacy protection: strengths and limitations of differential privacy. Monetary and Economic Studies **41**, 49–80 (2023)
16. N. Ponomareva, H. Hazimeh, A. Kurakin, Z. Xu, C. Denison, H.B. McMahan, S. Vassilvitskii, S. Chien, A.G. Thakurta, How to dp-fy ml: a practical guide to machine learning with differential privacy. Journal of Artificial Intelligence Research **77**, 1113–1201 (2023)
17. P.C.M. Arachchige, P. Bertok, I. Khalil, D. Liu, S. Camtepe, M. Atiquzzaman, Local differential privacy for deep learning. IEEE Internet Things J. **7**, 5827–5842 (2019)
18. E. Devaux. What is differential privacy: definition, mechanisms, and examples. https://www.anonos.com/blog/what-is-differential-privacy-definition-mechanisms-examples, accessed 4 July 2024. 2024.
19. B. Jiang, J. Li, G. Yue, H. Song, Differential privacy for industrial internet of things: opportunities, applications, and challenges. IEEE Internet Things J. **8**, 10430–10451 (2021)
20. M. Ivezic, L. Ivezic. Securing data labeling through differential privacy. https://defence.ai/ai-security/differential-privacy-ai/, accessed 4 July 2024. 2022.
21. V.V. Vegesna, Privacy-preserving techniques in AI-powered cyber security: challenges and opportunities. International Journal of Machine Learning for Sustainable Development **5**, 1–8 (2023)
22. M. Yang, T. Guo, T. Zhu, I. Tjuawinata, J. Zhao, K.-Y. Lam, Local differential privacy and its applications: a comprehensive survey. Computer Standards & Interfaces **89**, 103827 (2023)
23. T. Dhar, The California Consumer Privacy Act: The ethos, similarities and differences vis-a-vis the General Data Protection Regulation and the road ahead in light of California Privacy Rights Act. Journal of Data Protection & Privacy **4**, 170–192 (2021)
24. A. Mishra, Y.I. Alzoubi, M.J. Anwar, A.Q. Gill, Attributes impacting cybersecurity policy development: an evidence from seven nations. Comput. Secur. **120**, 102820 (2022)
25. A. Mishra, Y.I. Alzoubi, A.Q. Gill, M.J. Anwar, Cybersecurity enterprises policies: a comparative study. Sensors **22**, 538 (2022)
26. S. Muneer, U. Farooq, A. Athar, M. Ahsan Raza, T.M. Ghazal, S. Sakib, A critical review of artificial intelligence based approaches in intrusion detection: a comprehensive analysis. Journal of Engineering **2024**, 3909173 (2024)

27. WIPO. The interplay between privacy, machine learning and artificial intelligence. https://www.wipo.int/tech_trends/en/artificial_intelligence/ask_the_experts/techtrends_ai_lorica.html, accessed 5 July 2024. 2024.
28. P. Christiano. Differential privacy for secure machine learning In 2024. https://expertbeacon.com/differential-privacy-machine-learning/, accessed 5 July 2024. 2023.
29. A. Mathew, P. Amudha, S. Sivakumari. Deep learning techniques: an overview. In *Advanced machine learning technologies and applications. AMLTA 2020. Advances in intelligent systems and computing*, Hassanien, A., Bhatnagar, R., Darwish, A., Ed.; Springer, Singapore, 2021; Volume 1141, pp. 599–608.
30. H.C. Tanuwidjaja, R. Choi, S. Baek, K. Kim, Privacy-preserving deep learning on machine learning as a service—a comprehensive survey. Ieee Access **8**, 167425–167447 (2020)
31. Y.I. Alzoubi, A.E. Topcu, A.E. Erkaya, Machine learning-based text classification comparison: Turkish language context. Appl. Sci. **13**, 9428 (2023)
32. M. Massaro, J. Dumay, J. Guthrie, On the shoulders of giants: undertaking a structured literature review in accounting. Accounting, Auditing & Accountability Journal **29**, 767–801 (2016)
33. Y.I. Alzoubi, A. Mishra, Green blockchain–a move towards sustainability. J. Clean. Prod. **430**, 139541 (2023)
34. Y.I. Alzoubi, A.Q. Gill, A. Al-Ani, Empirical studies of geographically distributed agile development communication challenges: a systematic review. Information & Management **53**, 22–37 (2016)
35. M. Adnan, S. Kalra, J.C. Cresswell, G.W. Taylor, H.R. Tizhoosh, Federated learning and differential privacy for medical image analysis. Sci. Rep. **2022**, 12 (1953)
36. Y.I. Alzoubi, A. Al-Ahmad, A. Jaradat, V. Osmanaj, FOG Computing architecture, benefits, security, and privacy, for the internet of thing applications: an overview. J. Theor. Appl. Inf. Technol. **99**, 436–451 (2021)
37. C.M. Bowen, S. Garfinkel, Philosophy of differential privacy. Not. Am. Math. Soc. **68**, 1727–1739 (2021)
38. R. Cummings, D. Desfontaines, D. Evans, R. Geambasu, Y. Huang, M. Jagielski, P. Kairouz, G. Kamath, S. Oh, O. Ohrimenko. Advancing differential privacy: where we are now and future directions for real-world deployment. *arXiv preprint* arXiv:2304.06929 *2023*.
39. J. Domingo-Ferrer, D. Sánchez, A. Blanco-Justicia, The limits of differential privacy (and its misuse in data release and machine learning). Commun. ACM **64**, 33–35 (2021)
40. A. El Ouadrhiri, A. Abdelhadi, Differential privacy for deep and federated learning: a survey. IEEE access **10**, 22359–22380 (2022)
41. J. Feng, L.T. Yang, B. Ren, D. Zou, M. Dong, S. Zhang, Tensor recurrent neural network with differential privacy. IEEE Trans. Comput. **73**, 683–693 (2023)
42. G.S. Kumar, K. Premalatha, G.U. Maheshwari, P.R. Kanna, G. Vijaya, M. Nivaashini, Differential privacy scheme using Laplace mechanism and statistical method computation in deep neural network for privacy preservation. Eng. Appl. Artif. Intell. **128**, 107399 (2024)
43. N. Rodríguez-Barroso, G. Stipcich, D. Jiménez-López, J.A. Ruiz-Millán, E. Martínez-Cámara, G. González-Seco, M.V. Luzón, M.A. Veganzones, F. Herrera. Federated learning and differential privacy: software tools analysis, the Sherpa. ai FL framework and methodological guidelines for preserving data privacy. Inform. Fusion. **64**, 270–292 (2020)
44. J. Vasa, A. Thakkar, Deep learning: differential privacy preservation in the era of big data. Journal of Computer Information Systems **63**, 608–631 (2023)
45. K. Wei, J. Li, M. Ding, C. Ma, H.H. Yang, F. Farokhi, S. Jin, T.Q. Quek, H.V. Poor, Federated learning with differential privacy: algorithms and performance analysis. IEEE Trans. Inf. Forensics Secur. **15**, 3454–3469 (2020)
46. J. Zhang, Q. Huang, Y. Huang, Q. Ding, P.-W. Tsai, DP-TrajGAN: A privacy-aware trajectory generation model with differential privacy. Futur. Gener. Comput. Syst. **142**, 25–40 (2023)
47. Y. Zhao, J. Chen, A survey on differential privacy for unstructured data content. ACM Computing Surveys (CSUR) **54**, 1–28 (2022)
48. T. Zhu, D. Ye, W. Wang, W. Zhou, S.Y. Philip, More than privacy: applying differential privacy in key areas of artificial intelligence. IEEE Trans. Knowl. Data Eng. **34**, 2824–2843 (2020)
49. A. Ziller, D. Usynin, R. Braren, M. Makowski, D. Rueckert, G. Kaissis, Medical imaging deep learning with differential privacy. Sci. Rep. **11**, 13524 (2021)
50. P. Mangold, M. Perrot, A. Bellet, M. Tommasi. Differential privacy has bounded impact on fairness in classification. In Proceedings of the 40th International Conference on Machine Learning, PMLR. Honolulu, USA, 2023; pp. 23681–23705.
51. R. Rofougaran, S. Yoo, H-H. Tseng, S.Y-C. Chen. Federated quantum machine learning with differential privacy. In Proceedings of the 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE. Seoul, South Korea, 2024; pp. 9811–9815.
52. M. Senekane. Deployment of differential privacy for application in artificial intelligence. In Proceedings of the 2021 International Conference on Electrical, Computer and Energy Technologies (ICECET), IEEE. Cape Town, South Africa, 2021; pp. 1–3.
53. T. Zhu, S.Y. Philip. Applying differential privacy mechanism in artificial intelligence. In Proceedings of the 39th international conference on distributed computing systems (ICDCS), IEEE. Dallas, TX, USA, 2019; pp. 1601–1609.
54. BigID. Navigating AI data privacy: current hurdles, future paths. https://bigid.com/blog/navigating-ai-privacy/, accessed 5 July 2024. 2024.
55. D. Capellupo. How differential privacy can make your AI models more responsible. https://www.rtinsights.com/how-differential-privacy-can-make-your-ai-models-more-responsible/, accessed 2 July 2024. 2020.
56. C. Dilmegani. Differential privacy: how it works, benefits & use cases in 2024. https://research.aimultiple.com/differential-privacy/, accessed 3 July 2024. 2024.
57. ENCORA. What is differential privacy and how does it work? https://www.encora.com/insights/differential-privacy-what-is-it, accessed 3 July 2024. 2022.
58. S. Fathima. Using differential privacy to build secure models: tools, methods, best practices. https://neptune.ai/blog/using-differential-privacy-to-build-secure-models-tools-methods-best-practices, accessed 2 July 2024. 2023.
59. FractalAnalytics. Differential privacy in responsible AI. https://online.flippingbook.com/view/913296844/2/#zoom=true, accessed 2 July 2024. 2023.
60. F. Hartmann. Distributed differential privacy for federated learning. https://research.google/blog/distributed-differential-privacy-for-federated-learning/, accessed 3 July 2024. 2023.
61. NIST. NIST drafts privacy protection guidance for AI-driven research. https://www.healthcareitnews.com/news/nist-drafts-privacy-protection-guidance-ai-driven-research, accessed 1 July 2024. 2023.
62. PrivateAI. The basics of differential privacy & its applicability to NLU models. https://www.private-ai.com/en/2022/10/18/the-basics-of-differential-privacy-its-applicability-to-nlu-models/, accessed 2 July 2024. 2022.
63. Raun. Differential privacy and deep learning. https://www.geeksforgeeks.org/differential-privacy-and-deep-learning/, accessed 4 July 2024. 2023.
64. B. Roy. Differential privacy advances part 1: strengths & weaknesses. https://blog.openmined.org/differential-privacy-advances-part-1-strengths-weaknesses/, accessed 4 July 2024. 2023.
65. C. Wright, K. Rumsey. The strengths, weaknesses and promise of differential privacy as a privacy-protection framework. https://math.unm.edu/~knrumsey/pdfs/projects/DifferentialPrivacy.pdf, accessed 10 July 2024. 2020.
66. M.v. Rijmenam. Privacy in the age of AI: risks, challenges and solutions. https://www.thedigitalspeaker.com/privacy-age-ai-risks-challenges-solutions/, accessed 1 July 2024. 2023.

Publisher's Note